Data assimilation at UMCP: Local Ensemble Kalman Filtering and 4D EnKF

Department of Meteorology and the Chaos Group
University of Maryland

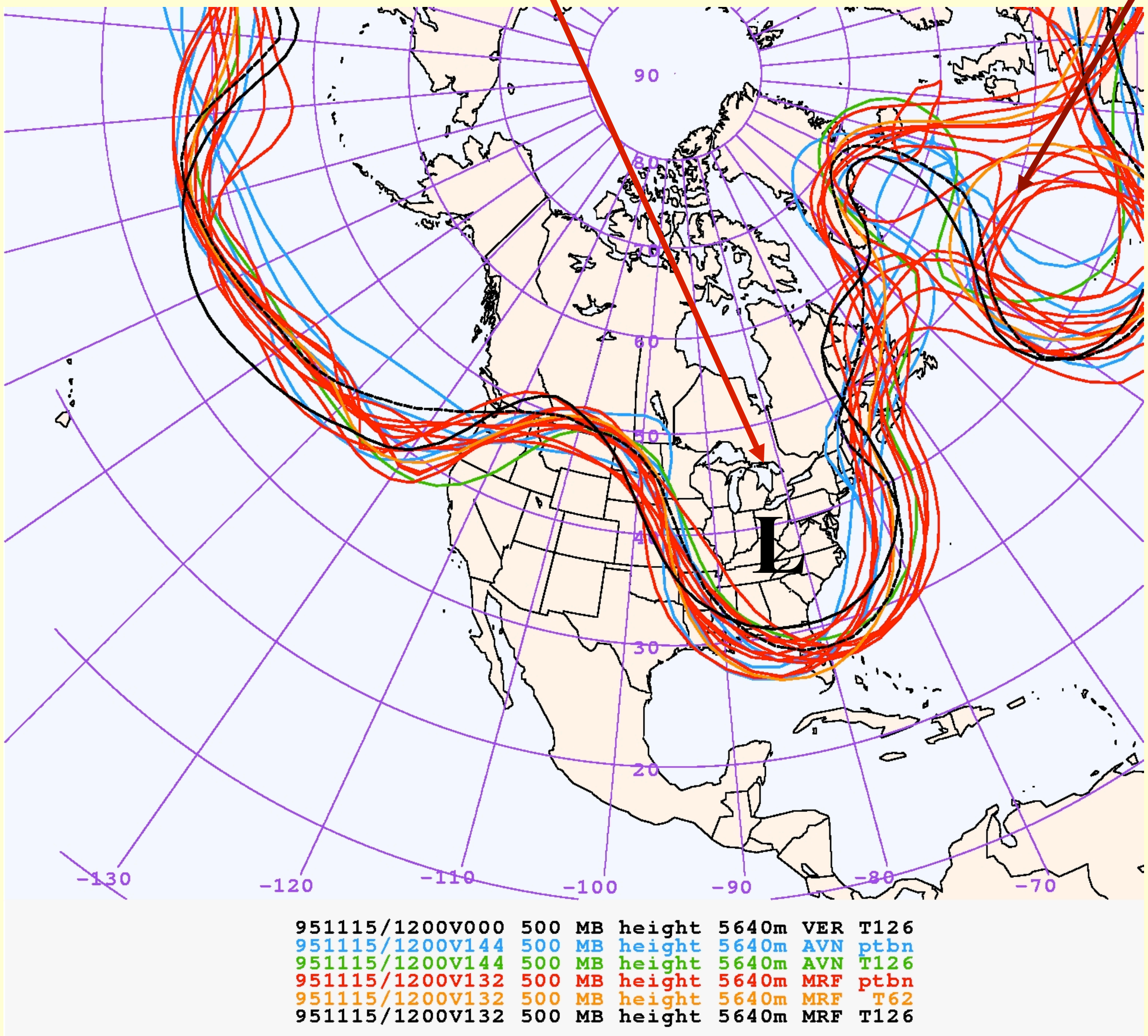
- Local Ensemble Kalman Filtering (LEKF)
- Examples of LEKF with a QG model, the Lorenz-40 model, and the NCEP global model
- (Extension to 4D Ensemble Kalman Filtering)
- Implications for AIRS data assimilation

The team

I. Szunyogh (UMCP), E. Kostelich (ASU), B. Hunt (UMCP), E. Ott (UMCP), E. Kalnay (UMCP), A. Zimin (UMCP), G. Gyarmati (UMCP), D. Patil (UMCP), J. Yorke (UMCP). Also, M. Corazza (Genoa), S. Yang (UMCP), T. Sauer (GMU)

An **interdisciplinary team** of experts in dynamical systems theory, meteorology, mathematics, and scientific computing

Example of a very predictable 6-day forecast, with “errors of the day”



Errors of the day tend to be localized and have simple shapes

(Patil et al, 2001)

Errors of the day

- They are instabilities of the background flow
- They dominate the analysis and forecast errors
- They are not taken into account in data assimilation except for 4D-Var and Kalman Filtering (very expensive methods)
- Our new Local Ensemble Kalman Filter includes the "errors of the day" but it can be done with present-day computers
- We have extended it to 4D EnKF (observations can be assimilated at the right time). No adjoint of the model required

3D-Var used in operational forecasting centers

$$J = \min \frac{1}{2}[(\mathbf{x}_b - \mathbf{x}_a)^{\mathrm{T}} \mathbf{B}^{-1} (\mathbf{x}_b - \mathbf{x}_a) + (\mathbf{y}_o - H\mathbf{x}_a)^{\mathrm{T}} \mathbf{R}^{-1} (y_o - Hx_a)]$$

Distance to forecast   Distance to observations

x is a model state vector, with $10^{6\text{-}8}$ d.o.f., and $\mathbf{y}_o$ is the set of observations, with $10^{5\text{-}9}$ d.o.f.
R is the observational error covariance,
B the forecast error covariance.

•In 3D-Var **B** is *assumed* to be constant: it does not include "errors of the day"
• 4D-Var is very expensive and does not provide the analysis error covariance.
• In Kalman Filtering **B** is forecasted. It is like running the model N times, where $N \sim 10^{6\text{-}8}$, so that it is impractical without simplifications

The solution: Ensemble Kalman Filtering

1) Perturbed observations and ensembles of data assimilation

- Evensen, 1994
- Houtekamer and Mitchell, 1998
- Keppene and Rienecker, 2002

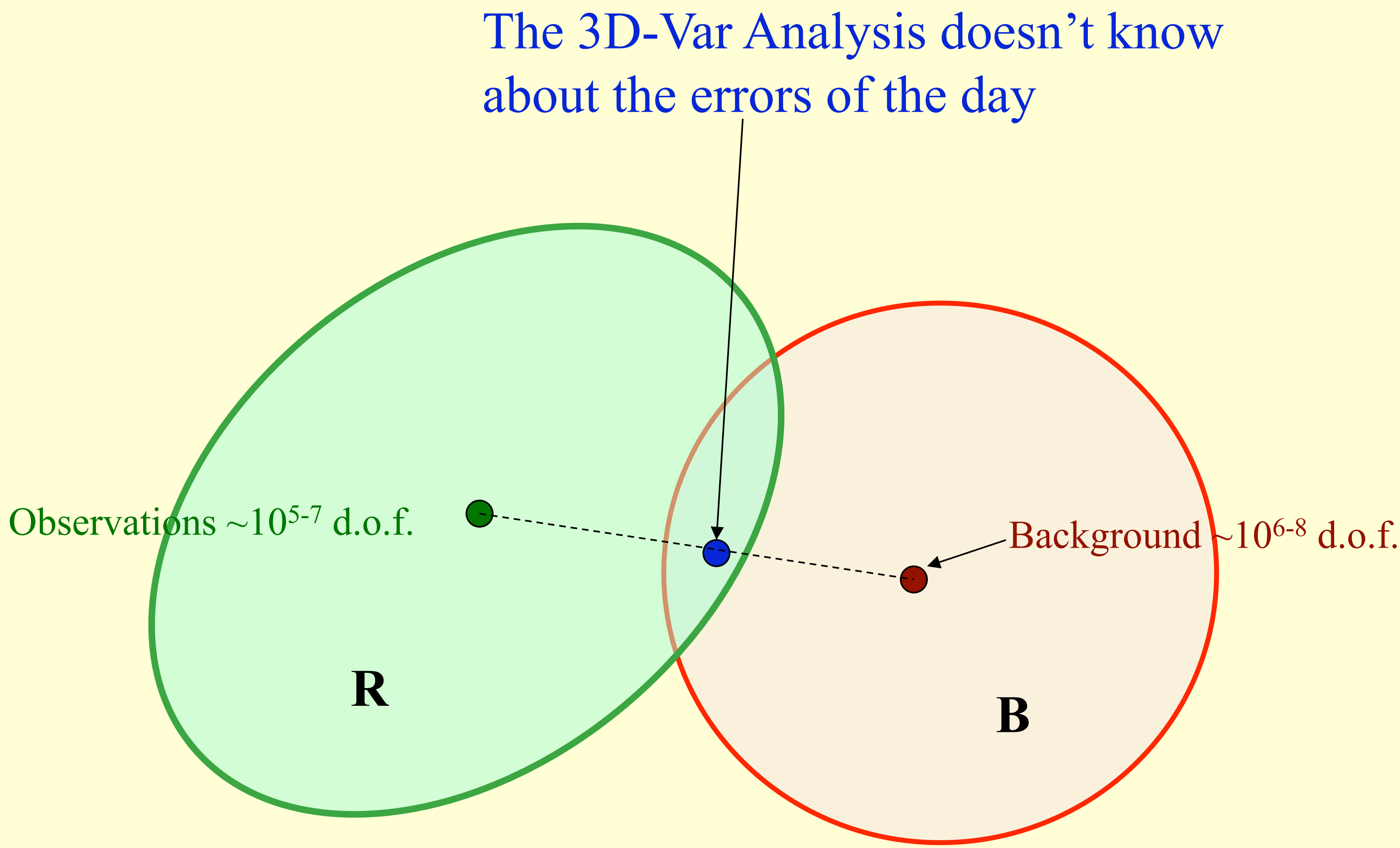
2) Square root filter, no need for perturbed observations:

- Tippett, Anderson, Bishop, Hamill, Whitaker, 2003
- Anderson, 2001
- Whitaker and Hamill, 2002
- Bishop, Etherton and Majumdar, 2001

3) Local Ensemble Kalman Filtering: done in local patches

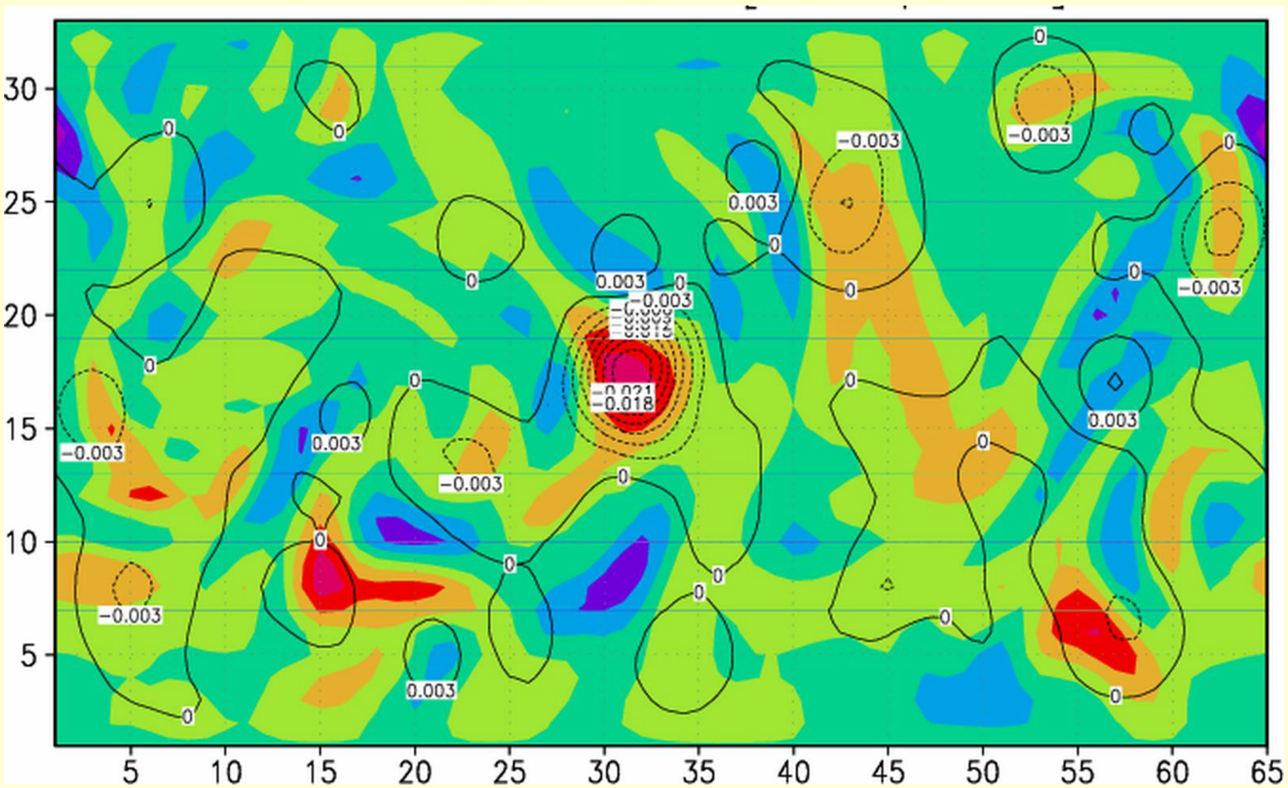
- Ott et al, 2003, MWR under review
- Hunt et al, 2003, Tellus: 4DEnKF

Suppose we have a 6hr forecast (background) and new observations



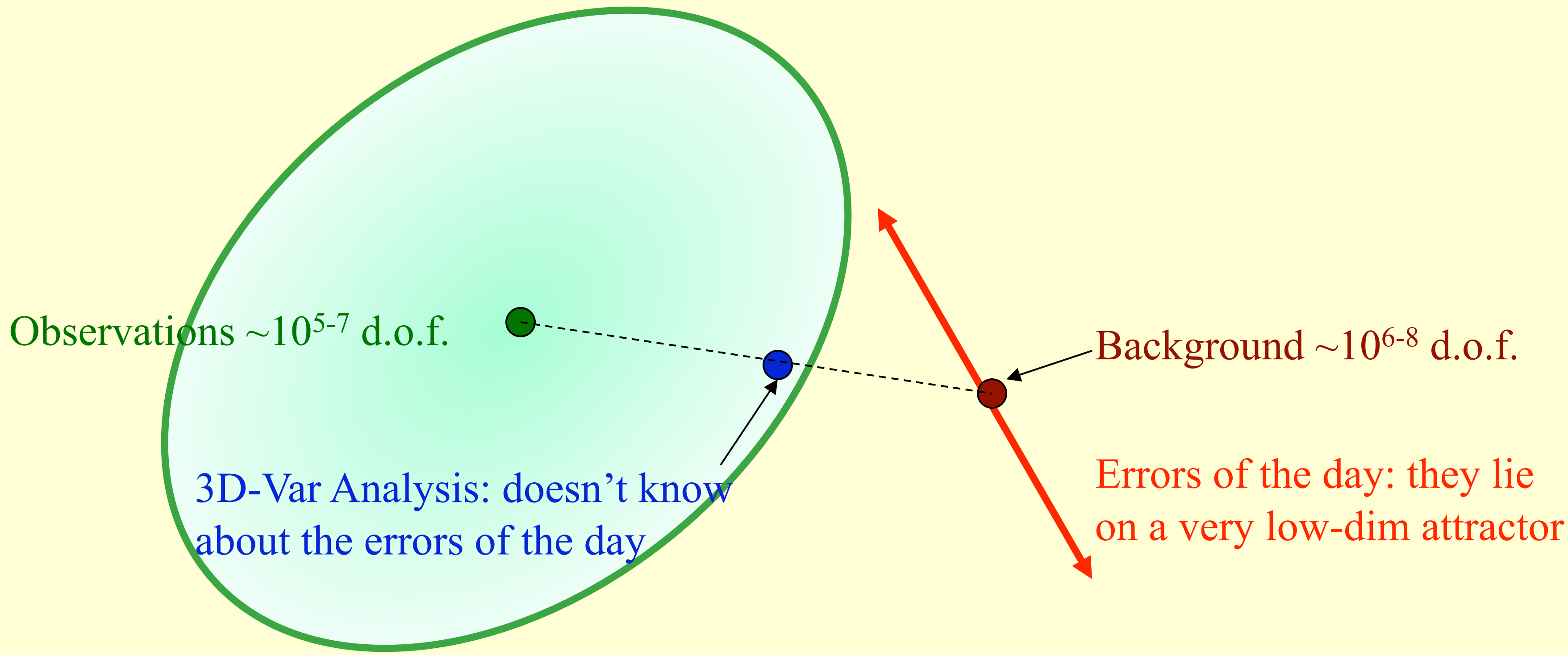
An example with a QG system (Corazza et al, 2003)

Background error (color) and 3D-Var analysis correction (contours)

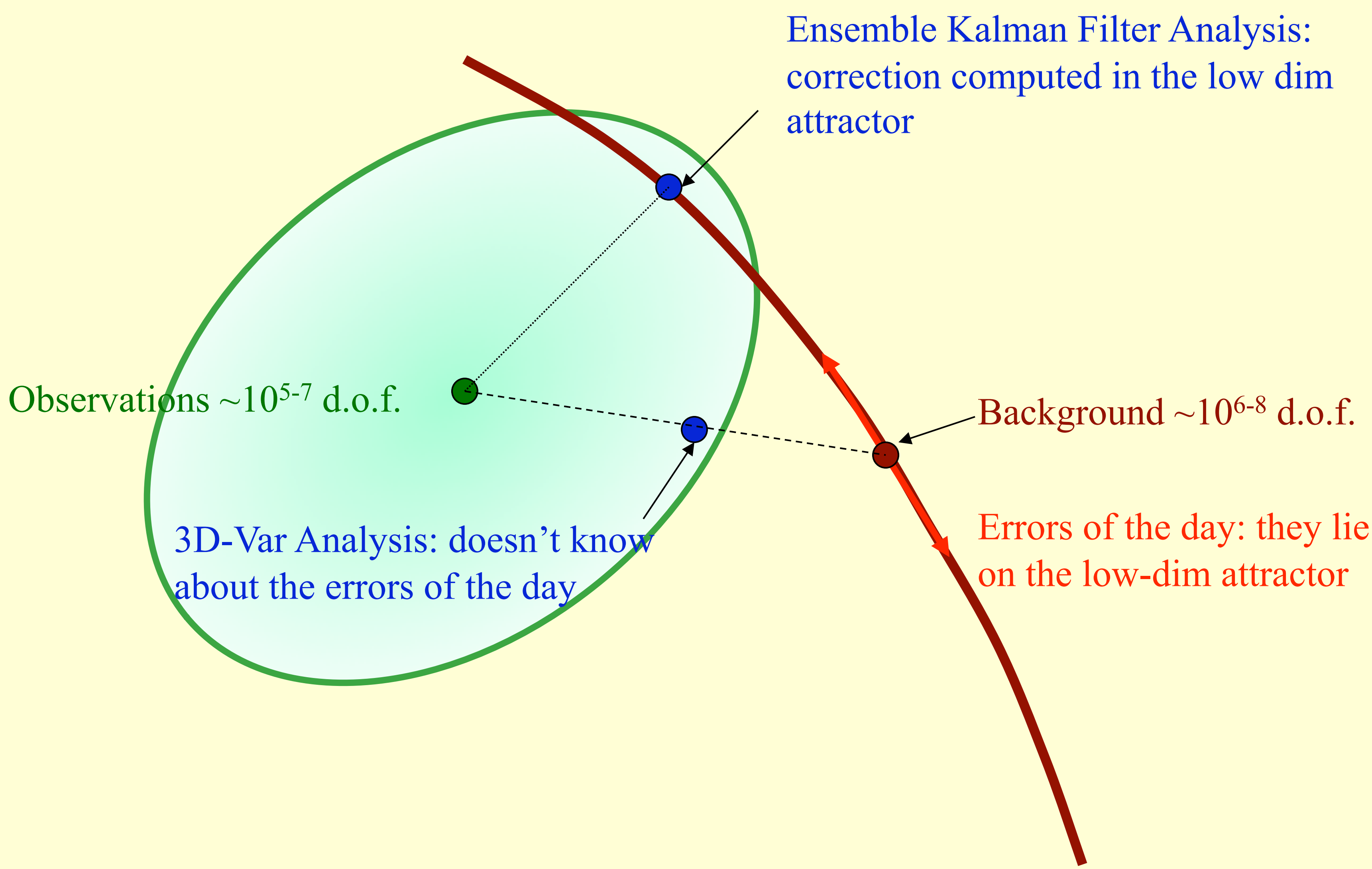


The analysis corrections due to the observations are isotropic because they don't know about the errors of the day

With Ensemble Kalman Filtering we get perturbations pointing to the directions of the "errors of the day"

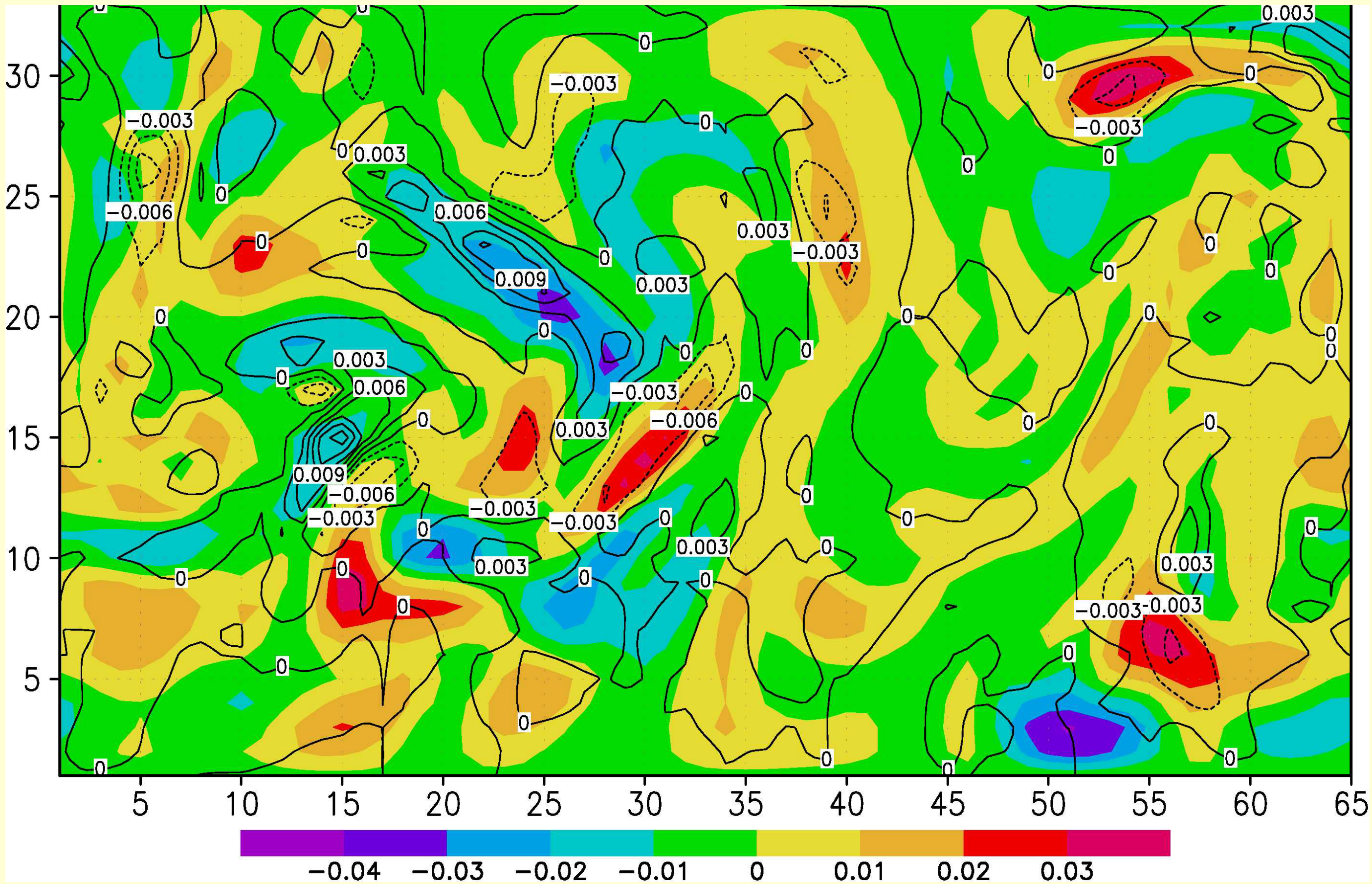


Ensemble Kalman Filtering is efficient because matrix operations are performed in the low-dimensional space of the ensemble perturbations



QG model example of Local Ensemble KF (Corazza et al)

Background error (color) and LEKF analysis correction

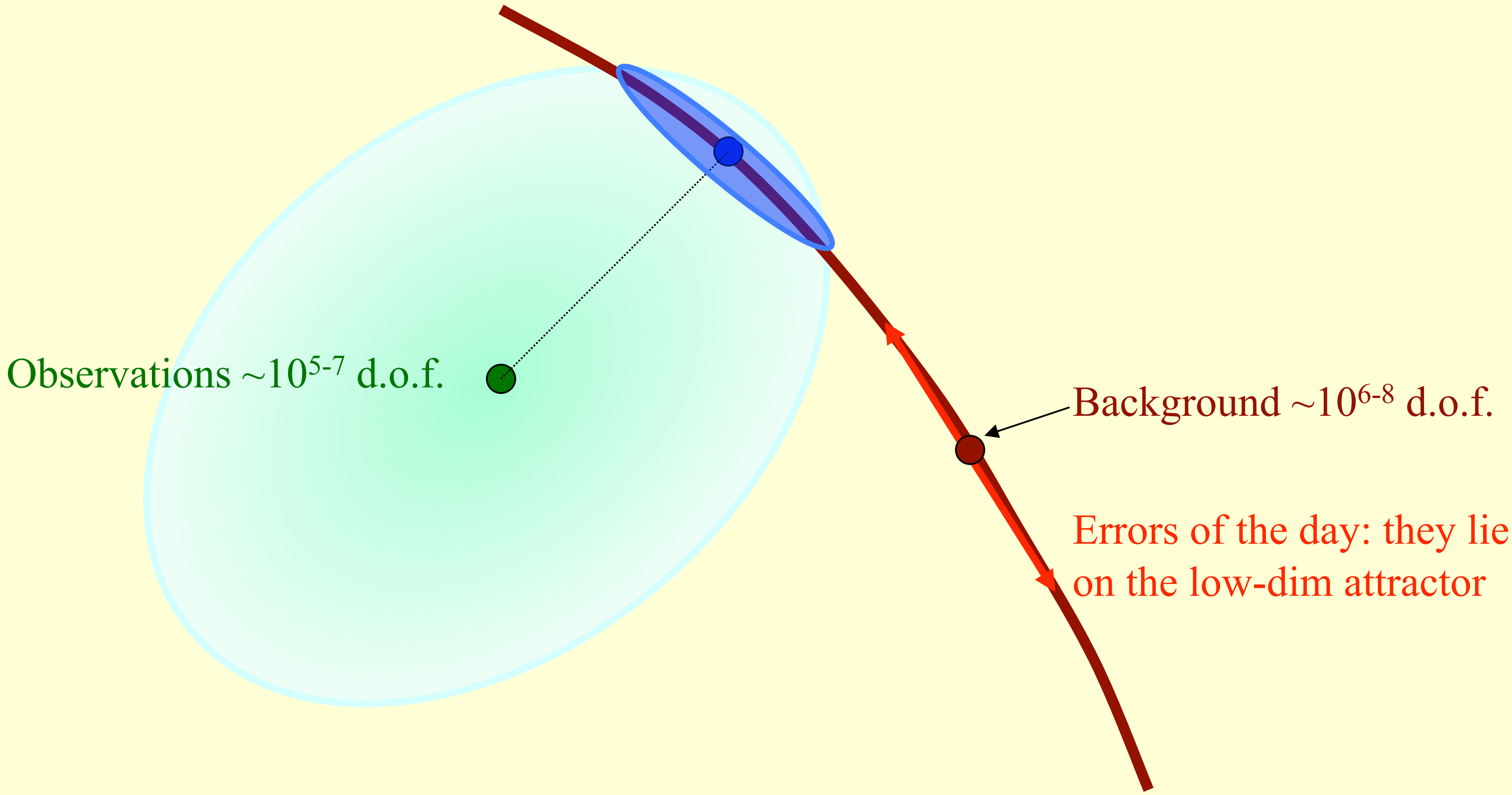


The LEKF does better because it captures the errors of the day

After the EnKF computes the analysis and the analysis error covariance $\mathbf{A}$, the new ensemble initial perturbations $\delta\mathbf{a}_i$ are computed:

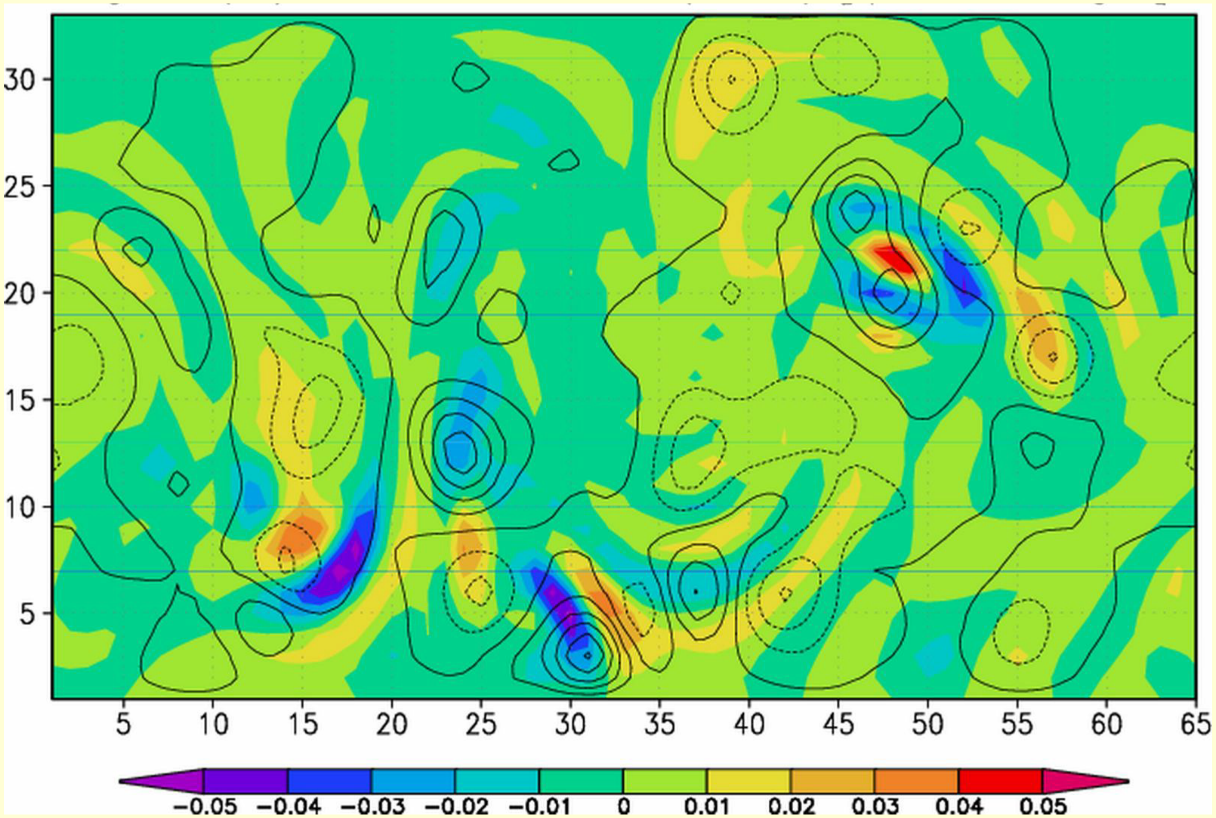
$$\sum_{i=1}^{k+1} \delta\mathbf{a}_i \delta\mathbf{a}_i^T = \mathbf{A}$$

These perturbations represent the analysis error covariance and are used as **initial perturbations** for the next ensemble forecast



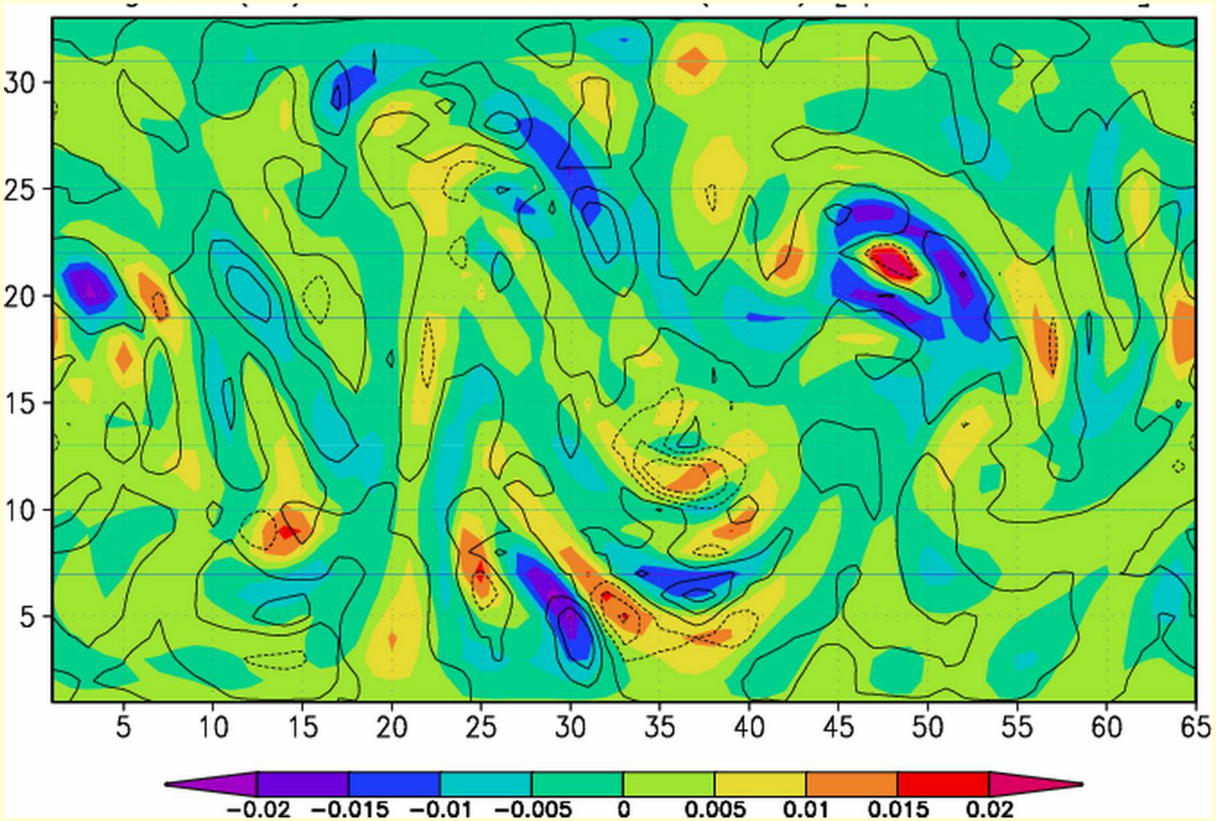
Again, from the QG simulation (Corazza et al, 2003)

Background error and 3D-Var analysis increment, June 15



The 3D-Var does not capture the errors of the day

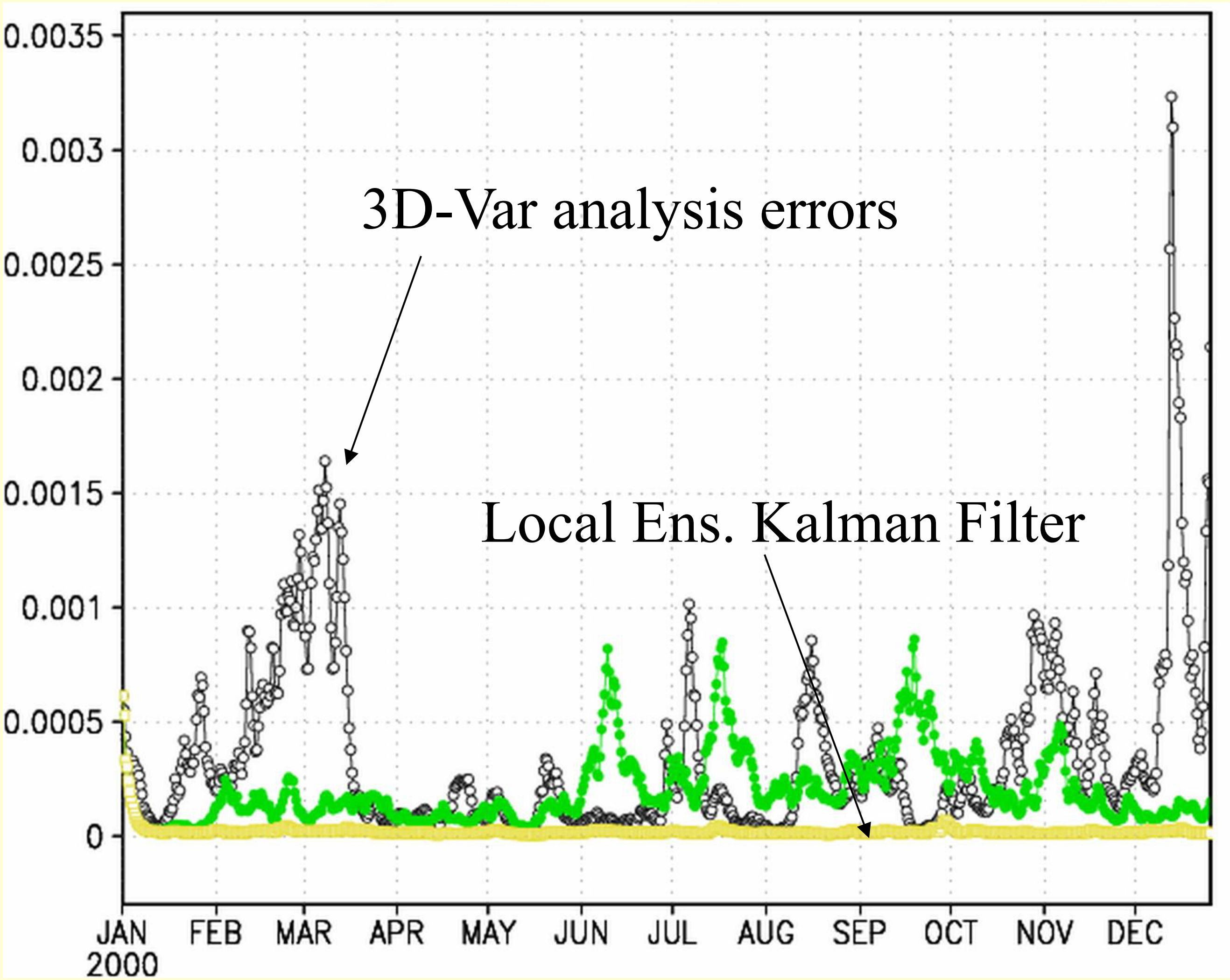
Background error (color) and LEKF analysis increments (contours), June 15



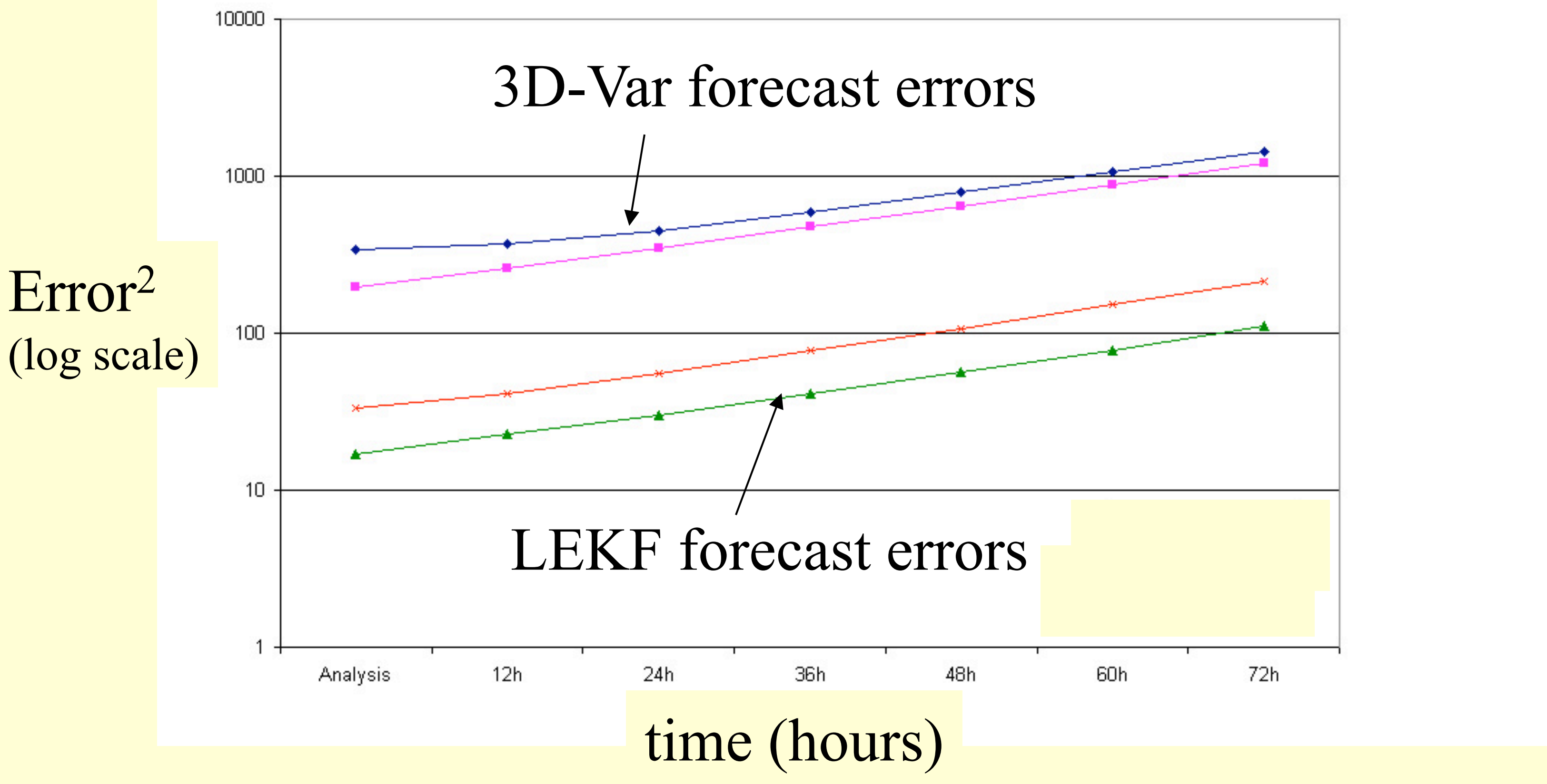
Contour interval: 0.005

The LEKF makes better use of the obs. because it includes the errors of the day

Area averaged Analysis Error: 3d-Var (black), LEKF (green), LEKF with covariance inflation (yellow)



This advantage continues into the 3-day forecasts



Why use a “local” ensemble approach?

• In the Local Ensemble Kalman Filter we compute the generalized “bred vectors” globally but use them **locally** (3D patches around each grid point of ~1000km x 1000km).

• These local patches (vertical columns) provide the **local** shape of the “errors of the day” for each grid point.

• At the end of the local analysis we create a new global analysis and initial perturbations from the solutions obtained at each grid point.

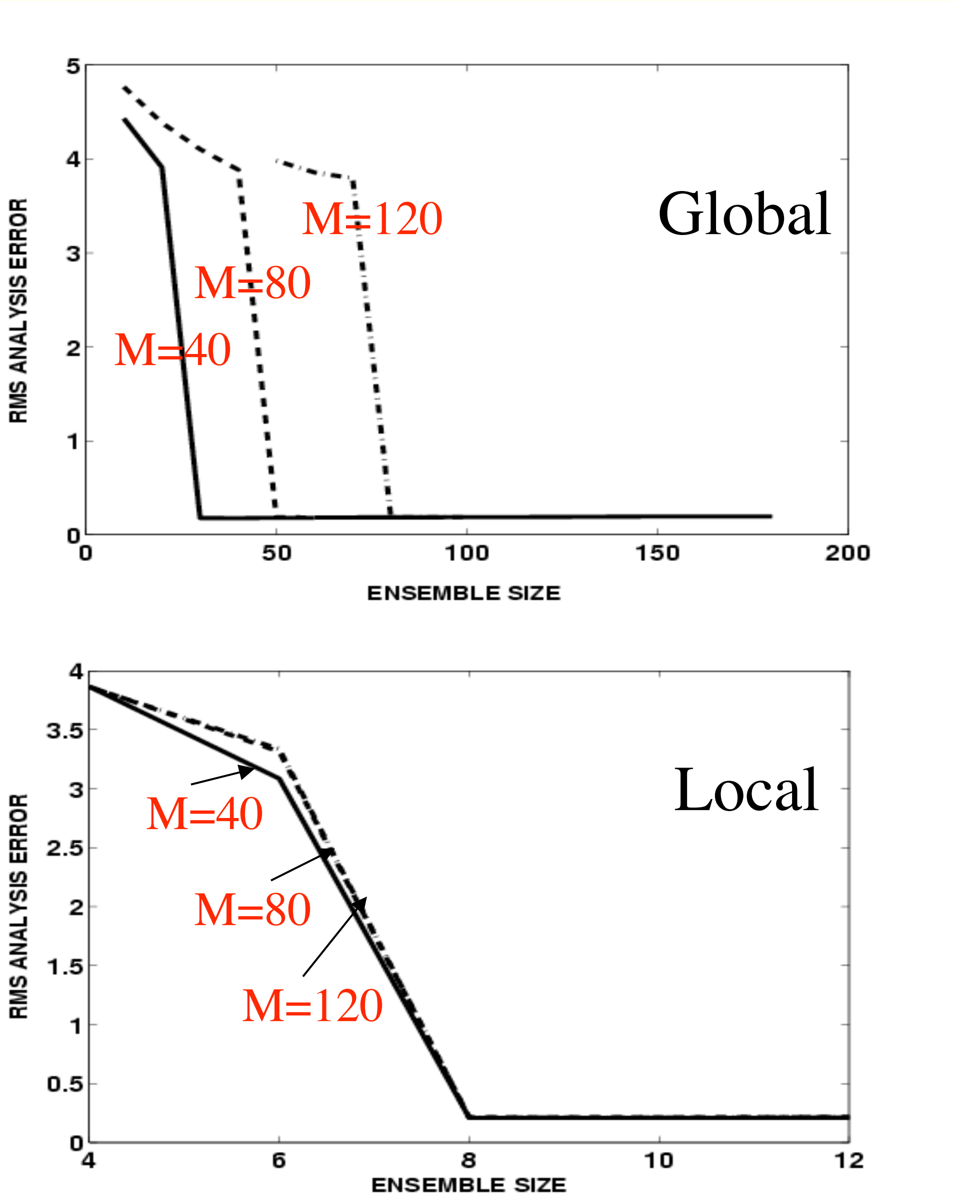
• **This reduces the number of ensemble members needed.**

• **It also allows to compute the KF analysis independently at each grid point (“embarrassingly parallel”)**.

Results with Lorenz 40 variable model

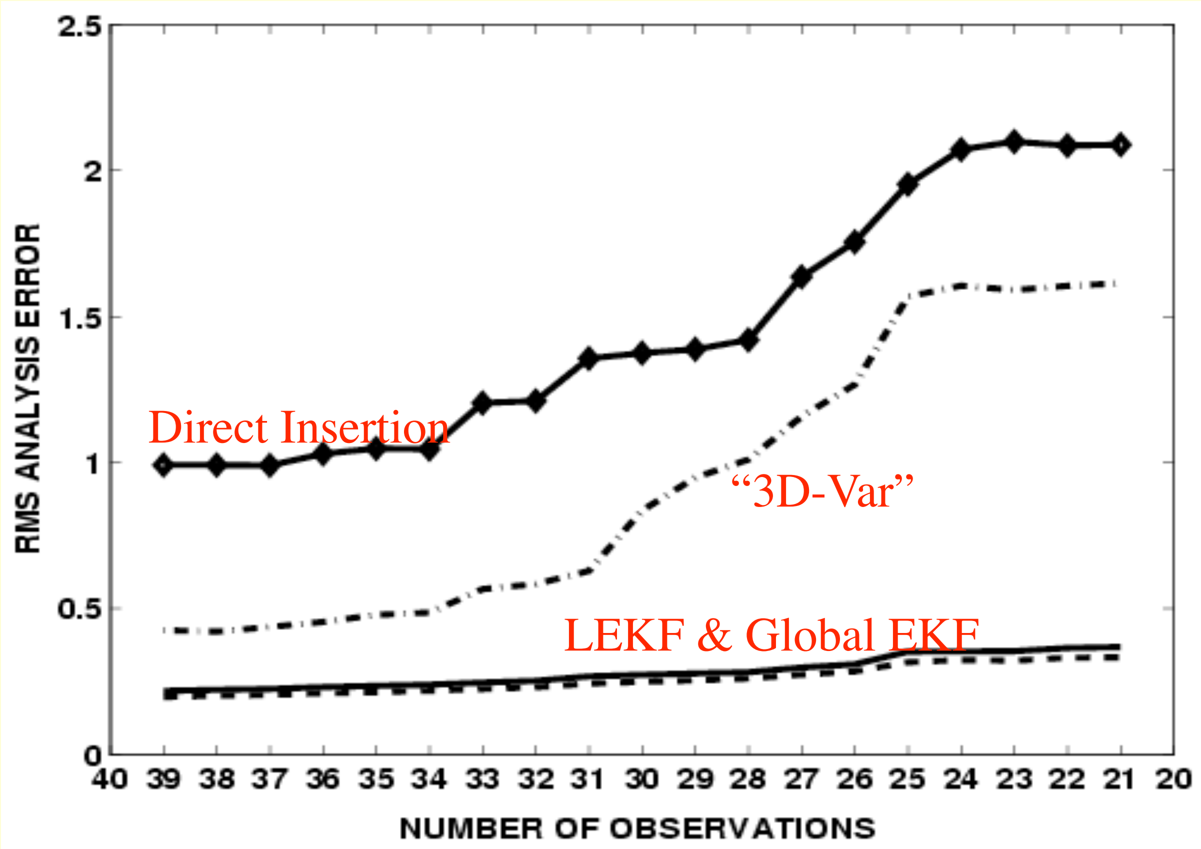
- Used by Anderson (2001), Whitaker and Hamill (2002) to validate their ensemble square root filter (EnSRF)
- A very large global ensemble Kalman Filter converges to an "optimal" analysis rms error=0.20
- This "optimal" rms error is achieved by the LEKF for a range of small ensemble members
- We performed experiments for different size models: M=40 (original), M=80 and M=120, and compared a global KF with the LEKF

LEKF vs. Global EKF



- The global scheme requires an increasing number of ensemble members as the size of the system increases
- The number of ensemble members needed in the LEKF is much lower than in the global scheme and independent of the system size

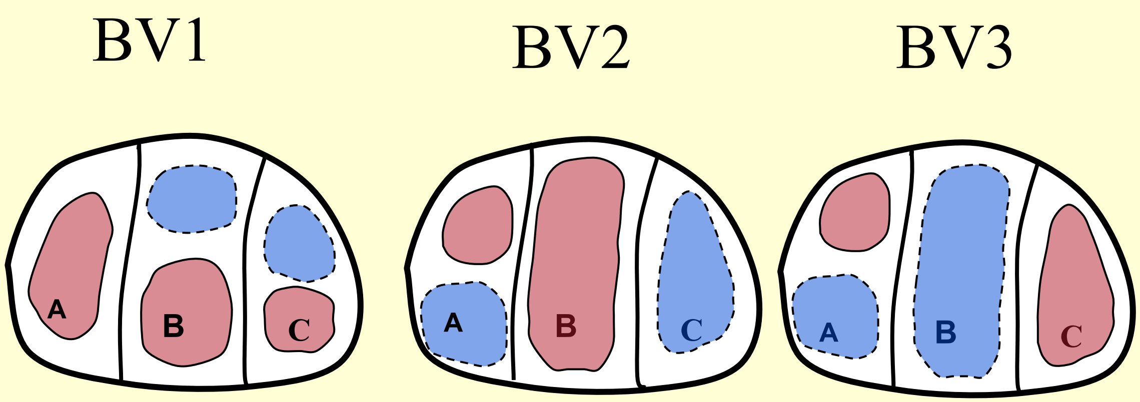
Reduced Observational Network



- The advantage of the Ensemble Kalman Filters increases as the number of observations decreases
- LEKF and Global EKF are similarly accurate independently of the number of observations

Why is the local analysis more efficient?

Schematic of a system with 3 independent regions of instability, A, B and C. Each region can have either wave #1 or #2 instability



From a local point of view, BV1 and BV2 are enough to represent **all** possible states.

From a global point of view, BV2 and BV3 are independent, and there are 63 possible different states…

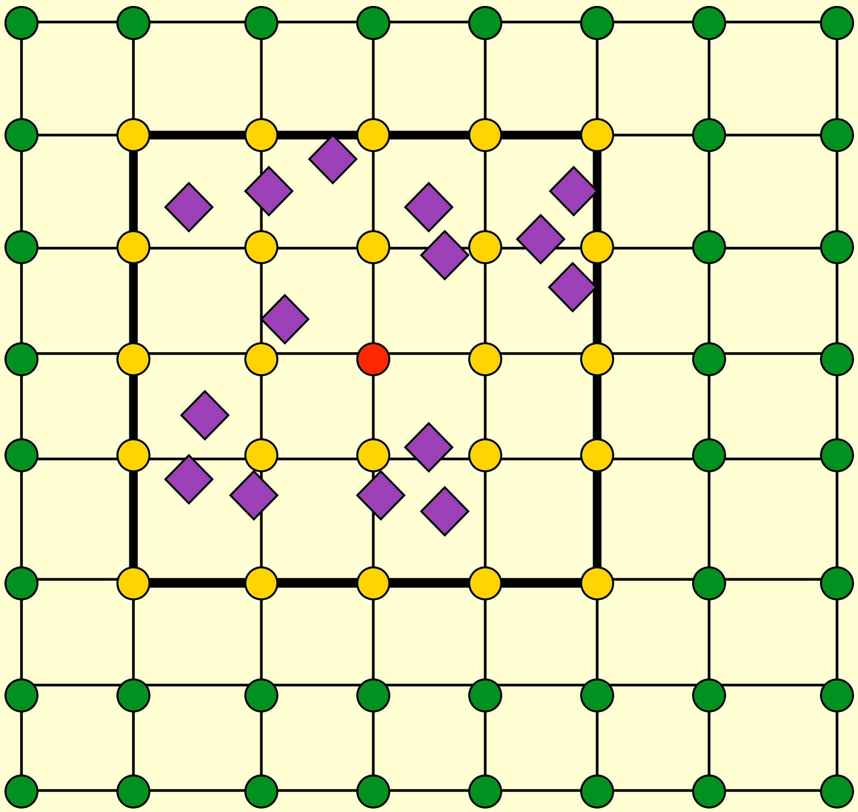
From Szunyogh, Kostelich et al

Preliminary LEKF results with NCEP's global model

- T62, 28 levels (1.5 million d.o.f.)
- The method is model independent: essentially the same code was used for the L40 model as for the NCEP global spectral model
- First simulation with observations at every grid point (1.5 million obs).
- Very parallel! Each grid point analysis is done independently
- Very fast! 20 minutes in a single 1GHz Intel processor with 10 ensemble members

Szunyogh et al.

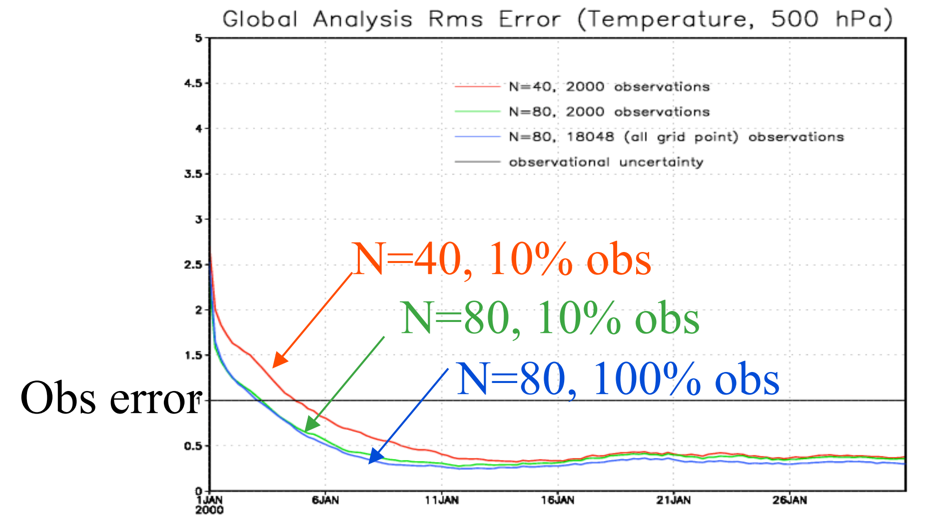
Local Ensemble Kalman Filter

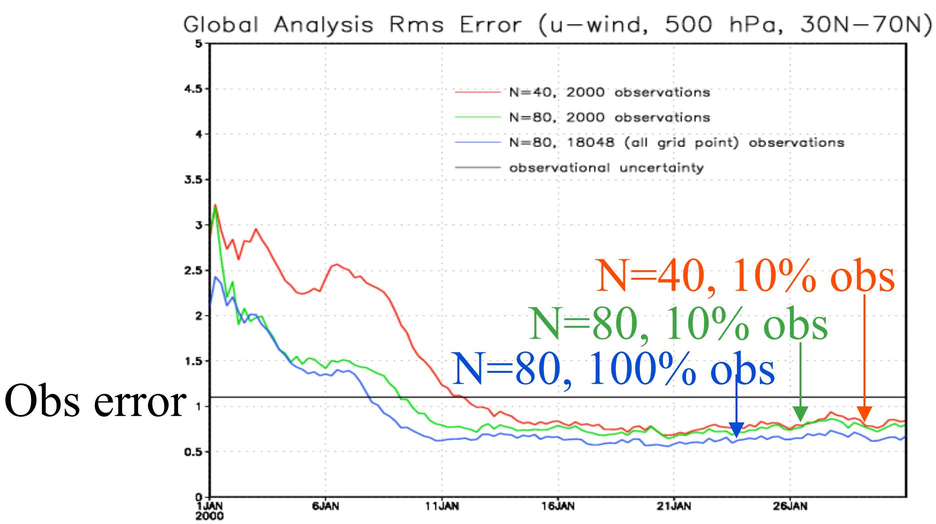


- A local column is defined **for each grid point**
- The KF**analysis** is **updated simultaneously** at the different grid-points
- **All observations** within the local region are **assimilated simultaneously**

Szunyogh et al.

Convergence of the LEKF





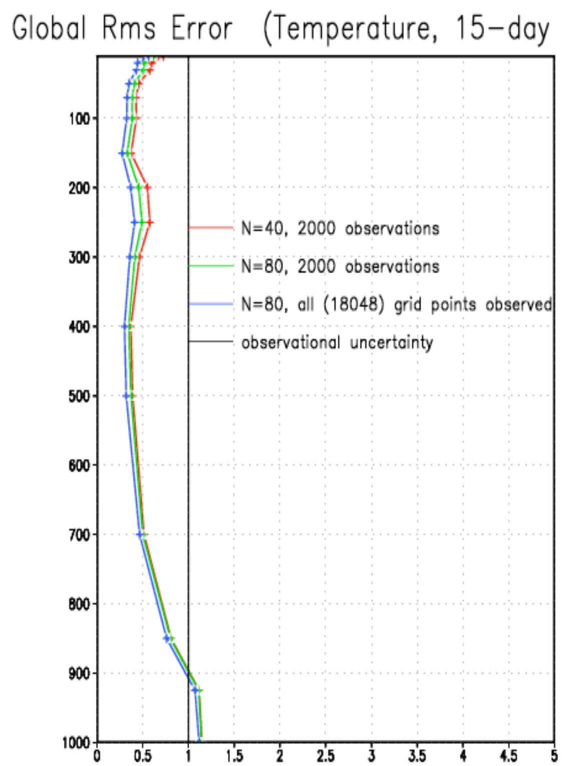
- For all observed variables, the analysis rms error converges to a value much smaller than the observational uncertainty
- For the wind components, the transient phase is longer, especially for the smaller ensembles
- For the wind components, the day-to-day variability is larger

Timing results

- Wind, temperature, and surface pressure observations at each grid point (a total of 1.5 million observations)
- 40 2.8 GH Xeon processors; 1 Gbit Ethernet for communication; parallel implementation (a $150,000 computer)
- Using 9x9x28 grid points patches, the entire data assimilation (including file movements, etc.) takes 6 minutes for a 40-member ensemble and about 12 minutes for an 80-member ensemble

Szunyogh at al.

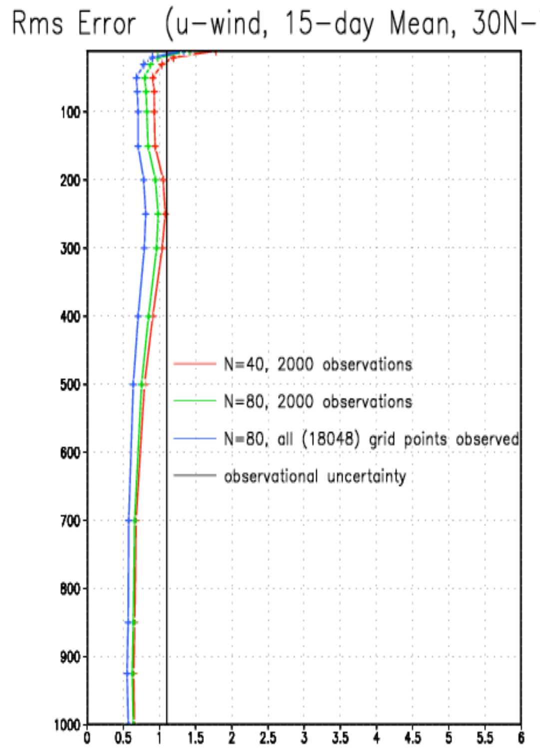
Temperature Rms Error Vertical Cross-Section



- Except for the boundary layer, the error is smaller than the observational uncertainty
- The relatively large global mean error in the boundary layer is due to large errors over land
- We believe this is because we use a local column, and will be improved by using several cubes

Szunyogh et al.

U-wind Rms Error Vertical Cross-Section



- The error is smaller than the observational uncertainty at all levels
- The errors are the largest in the jet layer
- The larger ensemble and better observational coverage have the most positive effect in the jet layer

Assessment of results

- The analysis errors are larger in layers of large vertical gradients (in the boundary layer for the temperature, and in the jet layer for the wind component)
- This suggests that a vertical localization is desirable (work in progress, local regions will be cubes of atmospheric air). It should also make it faster.
- We expect that this will also allow for the assimilation of humidity observations

Conclusions

- In the simulated observations tests, the LEKF is very fast, stable, and accurate.
- It can be extended to assimilate asynchronous observations at their right time at little additional cost (4DEnKF).
- It can assimilate all observations simultaneously.
- It can be implemented in current computers and provides “perfect ensemble perturbations”.
- The data assimilation code is model independent.
- Using cubes (rather than local columns) should improve the results and make LEKF even more efficient.
- We should be ready within a few months to use the LEKF with real observations.
- **LEKF with 4D EnKF is ideal for assimilating AIRS**

References and thanks:

Ott, Hunt, Szunyogh, Zimin, Kostelich, Corazza, Kalnay, Patil, Yorke, 2003: Local Ensemble Kalman Filtering, MWR, under review.

Patil, Hunt, Kalnay, Yorke and Ott, 2001: Local low-dimensionality of atmospheric dynamics, PRL.

Corazza, Kalnay, Patil, Yang, Hunt, Szunyogh, Yorke, 2003: Relationship between bred vectors and the errors of the day. NPG.

Hunt et al, 2003: 4DEnKF. Submitted to Tellus.

Szunyogh, I, E. Kostelich et al, 2003: Computational aspects of the implementation of LEKF. In preparation.

Kalnay, 2003: Atmospheric modeling, data assimilation and predictability, Cambridge University Press, 341 pp.